Continuous Sensing of Gesture for Control of Audio-Visual Media

Xin Wei Sha[1], Giovanni Iachello[2], Steven Dow[2], Yoichiro Serita[2], Tazama St. Julien[2], Julien Fistre[3]

[1] School of Literature, Communication, and Culture /
GVU Center
Georgia Institute of Technology
xinwei@lcc.gatech.edu

[2, 3] College of Computing/GVU Center
Georgia Institute of Technology
[2] {giac, steven, seri, stjulien}@cc.gatech.edu,
[3]gtg918b@ece.gatech.edu

Abstract

This note describes how continuous sensing of gesture enabling expressive control of real-time audio/visual media is achieved using Berkeley motes. We have contributed a relatively stable, inexpensive, extensible, and replicable wireless sensing platform for continuous motion tracking and placed the sensors into clothing to provide unobtrusive, natural affordances to the gesturing user. This paper includes an analysis of system requirements and a discussion of the software/hardware architecture.

1. Motivation

New interactors and embedded processors offer a wide range of control options for time-based media. We are interested in tackling the problem of mapping *freehand,* unclassified gestures to the real-time synthesis and modulation of video and sound, by using multidimensional generation models and leveraging the physical affordances and user expectations provided by cloth substrate or body placement. Such systems can be used for performance, play and entertainment environments.

Our user-centered design involves two sets of users: the players who perform in responsive media spaces and the designers of such spaces. For our application domain, we are aiming to support social settings with 3 or more players in a common space, to study elicited group social behaviors. For comfort and freedom of movement, it is essential that the people remain untethered, and that the physical devices be unobtrusive on the body or in clothing. For the other users, the designers, we have exposed a set of sensing functions to the Max/MSP/Jitter programming environment, because it is a *lingua franca* for professional live video/sound instrument designers and it allows the designer to rapidly implement plausible and tangible AV feedback in the media environment.

2. Requirements analysis and related work

Our wireless sensor platform has evolved over 5 generations, used in 10 experiments over the past 3 years, exploring a range of technical solutions [4]. In contrast with traditional wireless sensor networks [1, 3] the present architecture has the unique requirement for relatively high frequency (10-100Hz) sensor data. In trade-off we are not as concerned about the high power consumption needed for continuous transmission.

Three regimes of sensing are most relevant to our situation: (1) approximate (topological) location using magnetic and light fields, (2) contact based on point and strip Force Sensitive Resistors (FSR) and (3) follow-through gesture based on acceleration. Sustaining the illusion of continuous, direct interaction requires the system to support sensor readings with **tight latency and update frequency** bounds.

End-to-end latency is used to evaluate the responsiveness of interactive media and includes the time needed to sense changes actuated by the user, to modify its state, and to produce a feedback in the space. The maximum acceptable latency is determined by the human motor/perceptual system depending on interaction mode (visual, acoustic or haptic). Table 1 summarizes rough acceptable bounds in this situation, based on the Model Human Processor. Frequency is the number of sensors readings that occur per time period. To achieve high frequency (necessary for the illusion of continuity) and to support many concurrent sensors, wireless communication must use the available radio spectrum optimally. Besides low latency and high frequency, the following factors were also important in choosing the architecture:

- small size, solidity and ease of physical setup;
- simple interface to media generation engine;
- easy allocation of processing activities;
- inexpensive and easy to recreate;
- low power consumption.

After examining a number of options, we chose the Berkeley motes because they are small, inexpensive, and readily available wireless sensor modules. The motes platform is easy to modify for implementing non-standard sensors and it operates with the TinyOS, which has a broad user community.

Table 1 Human interaction bounds

Type of feedback	Frequency Hz	Latency ms
visual	15-25	50

acoustic	5-10	10-50
acoustic (melodic)	50	10
haptic	1000	1

3. Architecture and performance

Sensor data collected on the motes is processed and transmitted to a base station, and then forwarded to a host computer via a serial connection. There, the data is again processed and broadcast over the OSC protocol to the Max/MSP multimedia system, closing the feedback loop to the user. Two types of motes have been used for sensing: Rene motes a test bed for FSR sensors, and Mica motes for testing the performance of magnetometers, accelerometers and photometers. Custom software was written for both platforms, and the TinyOS networking and serial stacks were modified.

In this project, the FSR were placed in clothing along the sleeve to measure arm bending and in a glove to detect finger contact with objects. These sensors are lightweight, resistant, flexible, stable over time, and provide good dynamic range (~7 bits) and sensitivity. Accelerometers can detect the variation in the gravitational field when tilted, while magnetometers can sense relative variations in earth's field when rotated around the vertical axis, or when a magnet is within 10-20cm. Both kinds of sensors were positioned at the wrists and ankles to measure substantial body movements. The sensor installation was constrained by intrinsic orientation, hardware fragility and the mechanical stress of electrical wiring on the body.

To provide the best possible resolution, and thus support fine gesture nuance, active calibration software was implemented on the mote to leverage the sensor "sweet spot". Magnetometers need to be dynamically biased to provide reliable readings because their range is wider than the ADC input range. Moreover, each sensor (even of the same kind) has a characteristic DC value, noise energy and range. Readings from the sensors are thus individually filtered, scaled and the most significant 8 bits are kept. To increase robustness, all tuning parameters can be adjusted from the client application at run time by sending control messages to the mote.

Data is transferred in packets, each packet containing one sample for each sensor on the board. There is no routing, and the base station is assumed to be always in range (enhanced coil antennas were used to increase the range). To reduce congestion on the shared medium and increase performance, TinyOS 1.0 has been modified by:

– doubling the serial speed between base station and host to 38400bps;
– eliminating acknowledgement packets over the radio;
– reducing packet overhead from 7 to 2 bytes, keeping only an 8 bit destination address and 8 bit CRC.

After these changes, only 288 bits per packet are sent over the radio rather than 960, shortening the transmission delay significantly. Data is transmitted only when a sensor reading actually changes (push mode), or every second, to inform the host that the sensor is still alive.

Errors, congestion and delays cause packets to be dropped. In busy environments, radio errors can affect up to 20% of the total packets scheduled to be sent. This loss ratio is high if compared to wired networks, but is adequate for our application thanks to the type of feedback and adaptive nature of the human motor/perceptual system. Given the packet drop rate, stateful compression algorithms of sensor readings, like ADPCM, have been ruled out. Stateless compression such as µ-law was tested but provides insufficient resolution.

Table 2 Latency breakdown

Stage	Time (ms)	Notes
Sensor reading	3	Typically much less
MAC delay	7 (avg)	Random delay in [6,8]
Radio transm.	7.2	288 bits at 40kb/s
Serial transm.	2.1	80 bits at 38.4kb/s
Total	19.3	

End-to-end latency is dominated by sensing and feedback times. Table 2 shows a breakdown of the theoretical latencies in the system (actual measures have not yet been made). As for update frequency, 40Hz can be sustained with one mote and 25Hz with four. Referring to Table 1, these results are adequate to support tangible audio/visual feedback, but are not suited for controlling instruments meant for melodic music performance.

4. Conclusions

With our clothing sensor system we are now able to employ much more powerful digital media logic than previous attempts with continuous, on-body sensing and media synthesis. [2] Next we intend to extend evaluation relative to the designers using standard HCI techniques and to the players using subjective studies that probe notions of play, intentionality and causality.

5. References

[1] Bergbreiter S., et al., COTS-BOTS specifications, 2002. www.eecs.berkeley.edu/ ˜sbergbre/CotsBots/specs.pdf.

[2] Die Audio Gruppe, http:// www.audioballerinas.com, 2003.

[3] Mainwaring, A., et al., Wireless Sensor Networks for Habitat Monitoring in 1st ACM Workshop on Wireless Sensor Networks and Applications, 2002. Atlanta, GA, USA.

[4] Sha, X. et al., Final Report TGarden Project 869-2-2001 to the Daniel Langlois Foundation for Art, Science, and Technology, 2002.